

Taller 4

Santiago López Rodríguez, Manuel Alejandro Noriega Lizarazo y Xara Lucia Chamorro Aristizabal

Librerías

```
library("tidyverse")

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.0.6      v dplyr   1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library("naniar")
library("ggthemes")

## Warning: package 'ggthemes' was built under R version 4.0.5

library("readxl")
library("lubridate")

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

library("dplyr")
```

Punto 1

```
IHSM <- read_delim("Datos/IHSM.csv", delim = ";")
```

```
##
## -- Column specification -----
## cols(
##   paises_P = col_character(),
##   codigo = col_character(),
##   ano = col_double(),
##   ANS = col_double(),
##   DP = col_double(),
##   ESQ = col_double(),
##   BI = col_double(),
##   DA = col_double(),
##   AH = col_double(),
##   DR = col_double(),
##   SU = col_double()
## )
```

```
regresion_1 <- read_delim("Datos/Regresion_1.csv", delim = ";")
```

```
##
## -- Column specification -----
## cols(
##   ano = col_double(),
##   paises_P = col_character(),
##   GINId = col_double(),
##   IC = col_double(),
##   IDH = col_double()
## )
```

```
regresion_2 <- read_delim("Datos/Regresion_2.csv", delim = ";")
```

```
##
## -- Column specification -----
## cols(
##   ano = col_double(),
##   paises_P = col_character(),
##   GPS = col_double(),
##   GPE = col_double(),
##   GPD = col_double()
## )
```

```
Países <- read_excel("Datos/PAISES.xlsx")
```

Punto 3

```
IHSM <- left_join(IHSM, regresion_1, by = c("países_P", "ano"))
IHSM <- left_join(IHSM, regresion_2, by = c("países_P", "ano"))
IHSM <- left_join(IHSM, Países, by = "países_P")
```

Punto 4

```
IHSM <- IHSM %>%  
  mutate(across(.cols = c("países_P", "codigo", "IncomeGroup"), ~as.factor(.x)))
```

Punto 5

```
## Creamos la variable IHSM
```

```
# Consideramos la fiabilidad de las variables, con un indicador de suma
```

```
IHSM <- IHSM %>%  
  mutate(suma = ANS+DP+ESQ+BI+DA+AH+DR)
```

```
# Alfa de Cronbach
```

```
for (i in 4:10) {  
  dato <- var(IHSM[i], na.rm = TRUE)  
  print(dato)  
}
```

```
##          ANS  
## ANS 1.359364  
##          DP  
## DP 0.4287596  
##          ESQ  
## ESQ 0.00195891  
##          BI  
## BI 0.02935347  
##          DA  
## DA 0.0249627  
##          AH  
## AH 0.8264917  
##          DR  
## DR 3.404528
```

```
# Prueba
```

```
varianza_variables <- 1.363118+0.4301509+0.00195831+0.02944265+0.02500865+0.8292454+3.417732
```

```
varianza_indicador <- var(IHSM$suma, na.rm = TRUE)  
v <- 7
```

```
cronbach <- function(v, v1, v2){  
  cosa <- v/(v-1)  
  resto <- (v2-v1)/v2  
  print(cosa*resto)  
}
```

```
# Fiabilidad
```

```
cronbach(v, varianza_variables, varianza_indicador)
```

```
## [1] 0.6942663
```

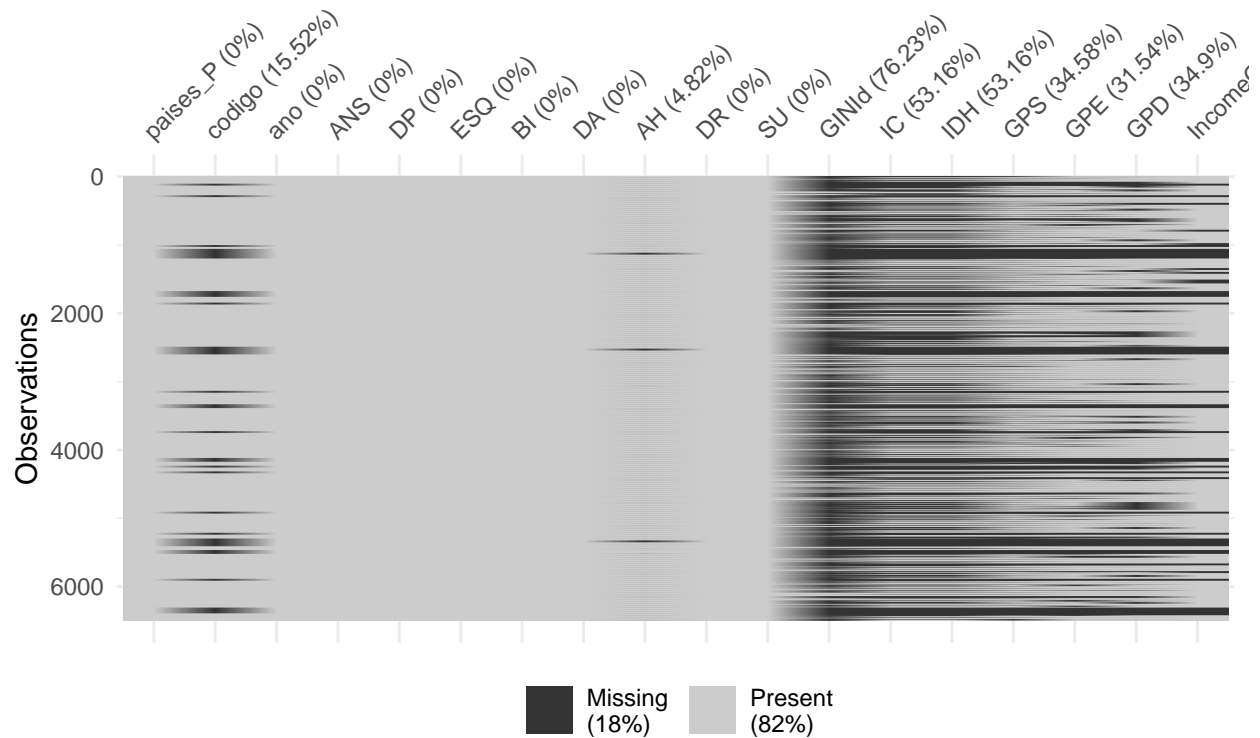
```
## la fiabilidad superior a 0.70 es lo preferible, como sale 0.6959472, se plantea como aceptable

rm(list= c("dato","regresion_1","regresion_2","i","v","varianza_variables","varianza_indicador"))

base_completa <- IHSM %>% select(-suma)

# Reemplazamos valores faltantes en las variables del indicador

# Solo AH contiene valores faltantes
vis_miss(base_completa)
```



```
# Reemplazo
base_completa <- base_completa %>%
  group_by(paises_P)

base_completa <- fill(base_completa, AH)

base_completa <- drop_na(base_completa, AH)

base_completa <- select(base_completa,-codigo)
```

Damos valores de normalizacion

```
for (i in 3:9) {  
  minimo<-min(base_completa[i])  
  print(minimo)  
}
```

```
## [1] 2.023393  
## [1] 2.139903  
## [1] 0.1469018  
## [1] 0.3145345  
## [1] 0.07390753  
## [1] 1.196984  
## [1] 9.715255
```

```
for (i in 3:9) {  
  maximo<-max(base_completa[i])  
  print(maximo)  
}
```

```
## [1] 8.96733  
## [1] 6.602754  
## [1] 0.3751096  
## [1] 1.206597  
## [1] 0.9439906  
## [1] 6.933015  
## [1] 19.11546
```

Indicador	Valor Maximo	Valor minimo
ANS	8.96733	2.023393
DP	6.602754	2.139903
ESQ	0.3751096	0.1469018
BI	1.206597	0.3145345
DA	0.9439906	0.07390753
AH	6.933015	1.196984
DR	19.11546	9.715255

#Calculamos los sub-indices

```
base_completa <- base_completa %>%  
  mutate(ANS_a = (ANS-2.023393)/(8.96733-2.023393),  
         DP_a = (DP-2.139903)/(6.602754-2.139903),  
         ESQ_a = (ESQ-0.1469018)/(0.3751096-0.1469018),  
         BI_a = (BI-0.3145345)/(1.206597-0.3145345),  
         DA_a = (DA-0.07390753)/(0.9439906-0.07390753),  
         AH_a = (AH-1.196984)/(6.933015-1.196984),  
         DR_a = (DR-9.715255)/(19.11546-9.715255))  
base_completa <- base_completa %>%  
  mutate(IHSM = ANS_a*1/7+DP_a*1/7+ESQ_a*1/7+BI_a*1/7+DA_a*1/7+AH_a*1/7+DR_a*1/7)
```

```
arrange(base_completa, desc(IHSM))
```

```
## # A tibble: 6,412 x 25
## # Groups:   paises_P [228]
##   paises_P  ano  ANS  DP  ESQ  BI  DA  AH  DR  SU  GINId  IC
##   <fct>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Austral~ 2007  6.67  4.91 0.367 1.15 0.907 3.63 18.7 10.5 NA      86
## 2 Austral~ 2006  6.68  4.92 0.367 1.15 0.896 3.64 18.7 10.3 NA      87
## 3 Austral~ 2008  6.65  4.88 0.366 1.15 0.916 3.62 18.7 10.7 35.4  80
## 4 Austral~ 2009  6.63  4.86 0.366 1.15 0.923 3.62 18.6 10.7 NA      87
## 5 Austral~ 2005  6.68  4.93 0.367 1.15 0.884 3.64 18.7 10.4 NA      88
## 6 Austral~ 2011  6.62  4.81 0.365 1.15 0.932 3.64 18.6 10.6 NA      88
## 7 Austral~ 2012  6.61  4.79 0.365 1.15 0.936 3.68 18.6 10.5 NA      85
## 8 Austral~ 2010  6.62  4.83 0.366 1.15 0.928 3.61 18.6 10.8 34.7  87
## 9 Austral~ 2013  6.61  4.76 0.365 1.15 0.939 3.71 18.5 10.5 NA      81
## 10 Austral~ 2014  6.60  4.73 0.365 1.15 0.942 3.74 18.5 11.0 34.4  80
## # ... with 6,402 more rows, and 13 more variables: IDH <dbl>, GPS <dbl>,
## #   GPE <dbl>, GPD <dbl>, IncomeGroup <fct>, ANS_a <dbl>, DP_a <dbl>,
## #   ESQ_a <dbl>, BI_a <dbl>, DA_a <dbl>, AH_a <dbl>, DR_a <dbl>, IHSM <dbl>
```

```
base_completa <- base_completa %>%
  mutate(IHSM = IHSM*100)
```

Estadística Descriptiva

```
resumen_paises <- base_completa %>%
  summarize(across(.cols = c("IHSM", "GINId", "IC", "IDH", "GPS", "GPE", "GPD"), list(Media=~mean(.x, na.rm = T),
```

```
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
## Warning in max(.x, na.rm = T): ningun argumento finito para max; retornando -Inf
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

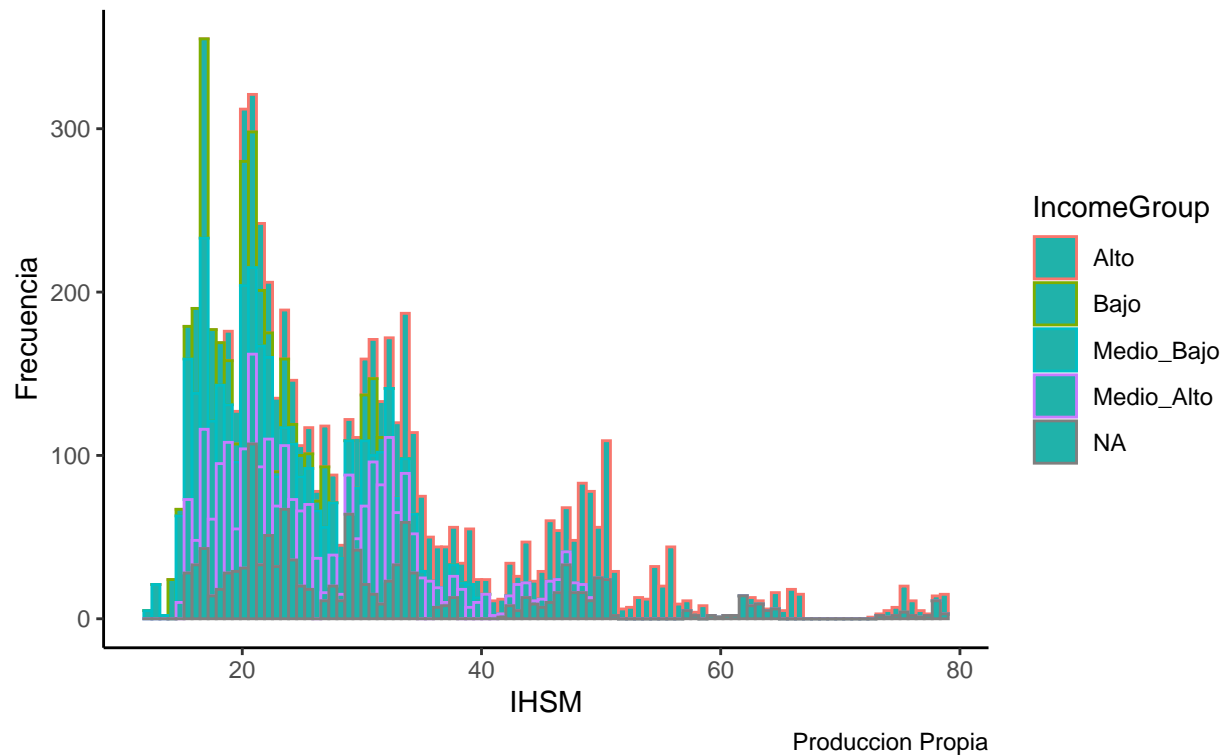
Punto 6

```
# Histogramas de la variable dependiente, e independientes principales
base_completa <- base_completa %>%
  mutate(IncomeGroup = recode(IncomeGroup,"High income"= "Alto","Upper middle income" = "Medio_Alto","Low income" = "Bajo"))

## IHSM
ggplot(data = base_completa,
       mapping = aes(x= IHSM,color = IncomeGroup)) +
  geom_histogram( fill = "lightseagreen",bins = 100)+
  labs(title = "Histograma de IHSM", subtitle = "Índice Humano de Salud Mental",y = "Frecuencia" , caption = "Fuente: Encuesta de Salud Mental de Chile 2019")
theme classic()
```

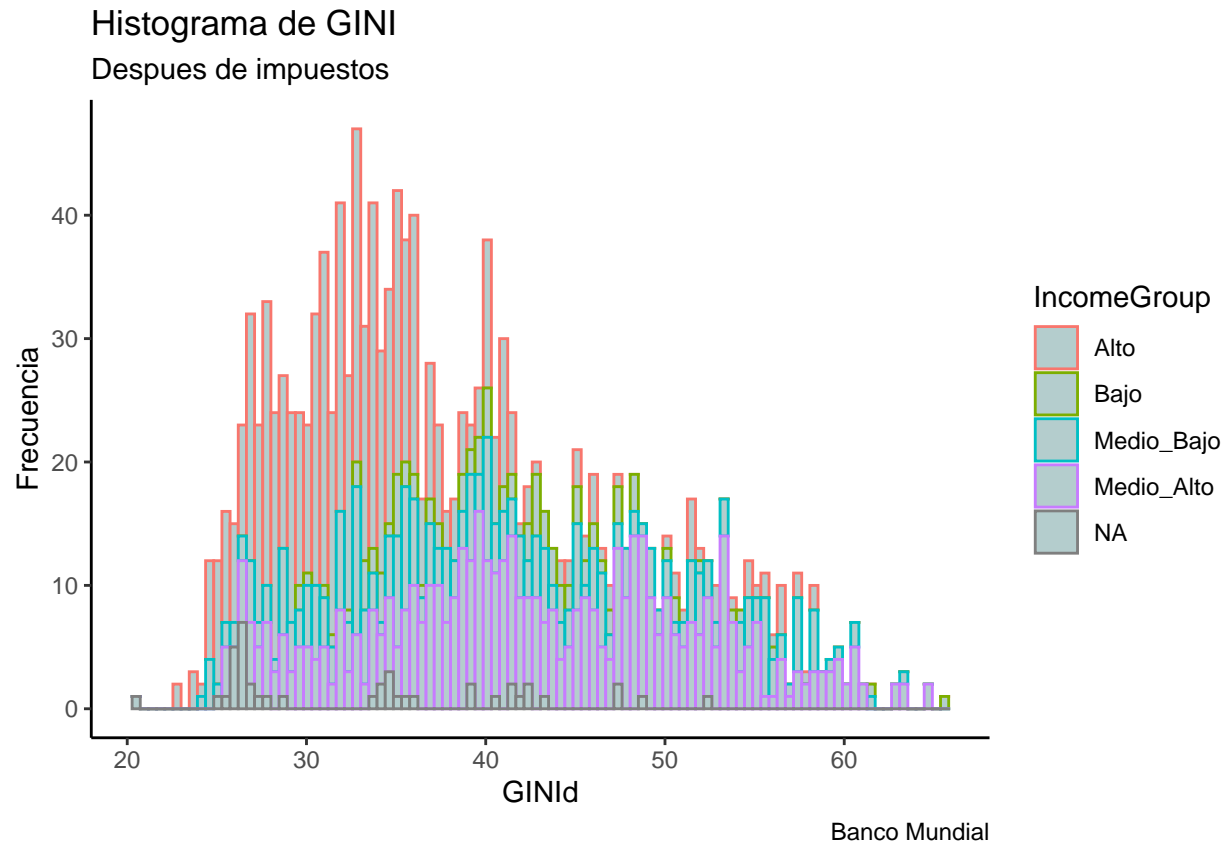
Histograma de IHSM

Índice Humano de Salud Mental



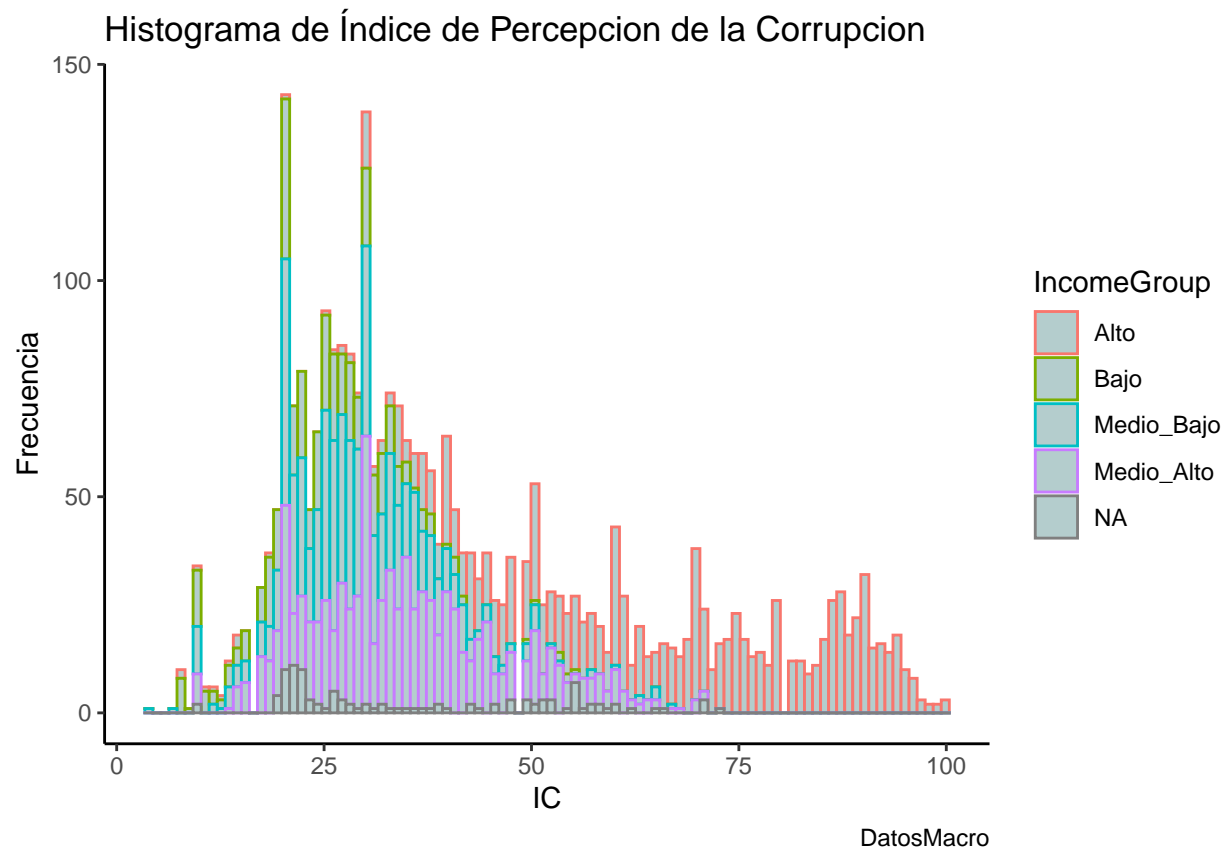
```
##GINId
ggplot(data = base_completa,
       mapping = aes(x= GINId,color = IncomeGroup)) +
  geom_histogram(fill = "lightcyan3",bins = 100)+
  labs(title = "Histograma de GINI", subtitle = "Despues de impuestos",y = "Frecuencia" , caption = " ")
  theme_classic()
```

```
## Warning: Removed 4868 rows containing non-finite values (stat_bin).
```



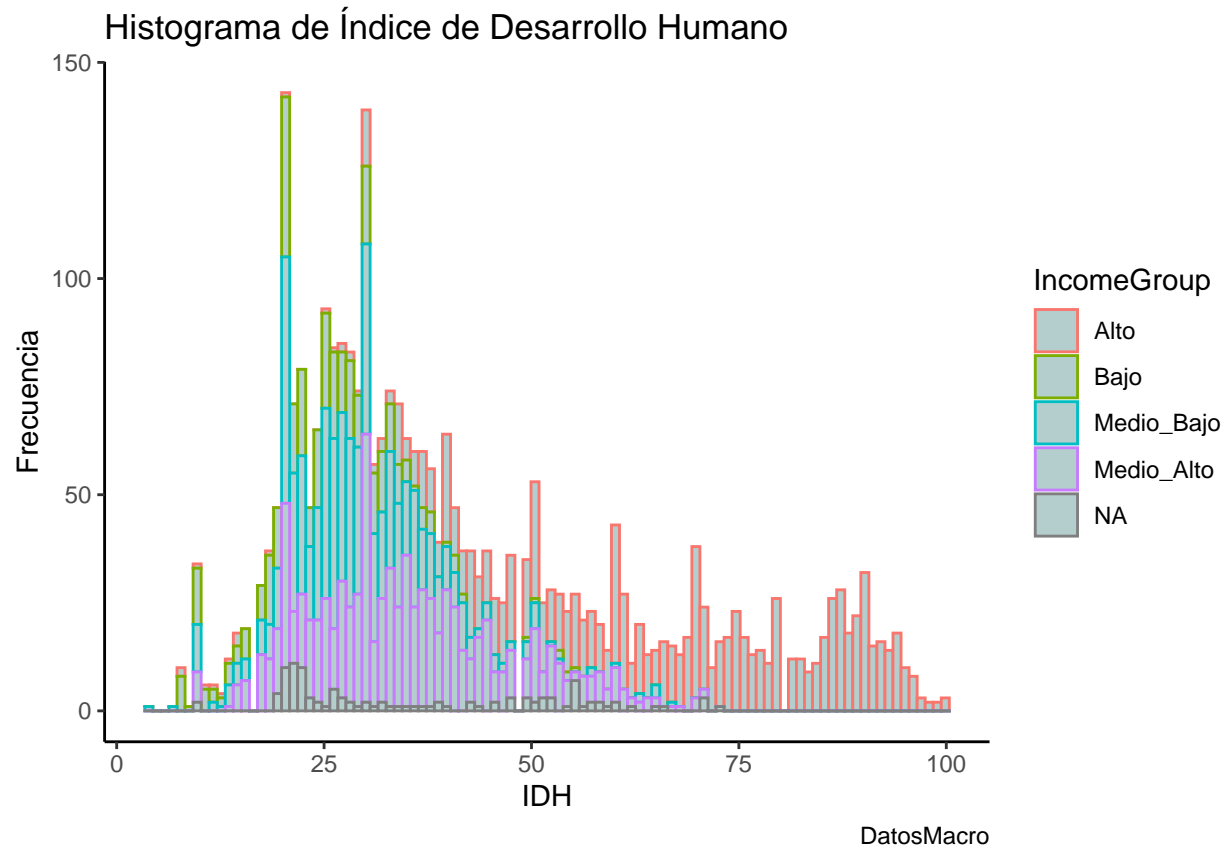
```
## IC
ggplot(data = base_completa,
       mapping = aes(x= IC,color = IncomeGroup)) +
  geom_histogram(fill = "lightcyan3",bins = 100)+
  labs(title = "Histograma de Índice de Percepcion de la Corrupcion",y = "Frecuencia" , caption = "Data")
  theme_classic()
```

```
## Warning: Removed 3369 rows containing non-finite values (stat_bin).
```



```
## IDH
ggplot(data = base_completa,
       mapping = aes(x= IDH,color = IncomeGroup)) +
  geom_histogram(fill = "lightcyan3",bins = 100)+
  labs(title = "Histograma de Índice de Desarrollo Humano",y = "Frecuencia", caption = "DatosMacro")+
  theme_classic()
```

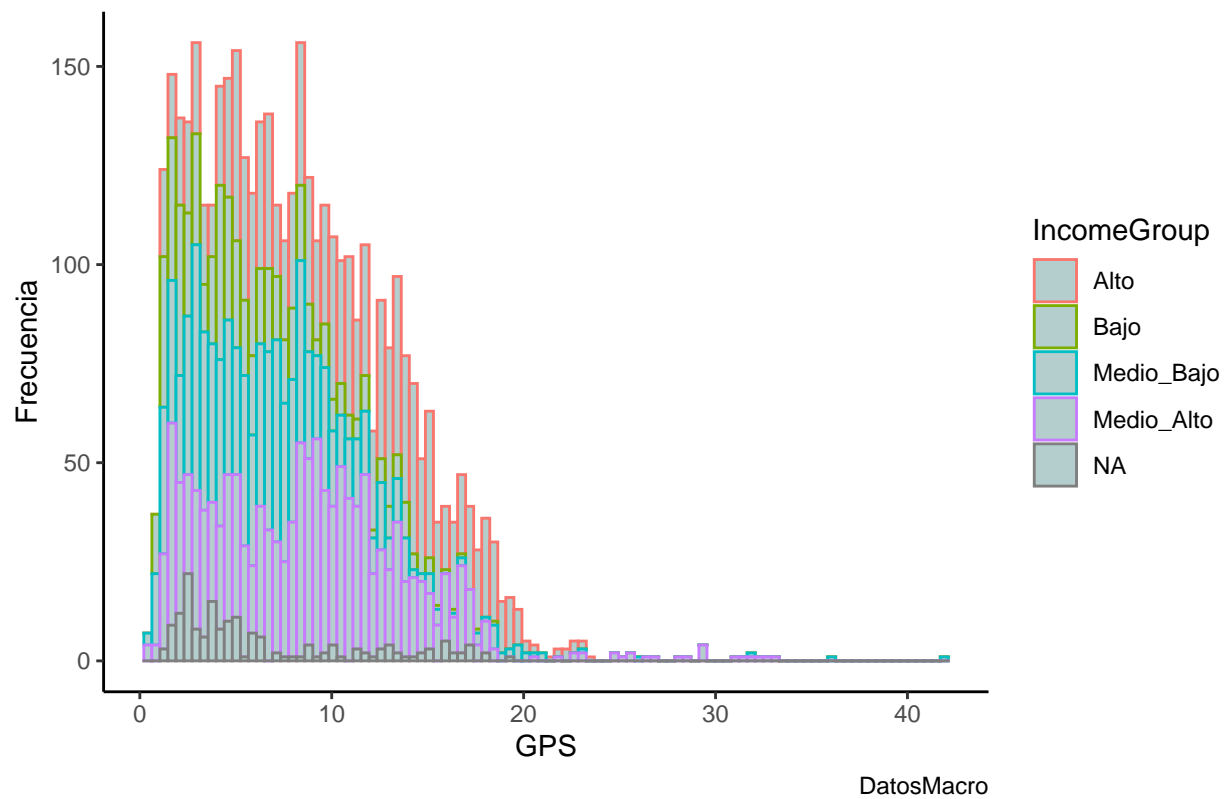
```
## Warning: Removed 3369 rows containing non-finite values (stat_bin).
```

```
## GPS
ggplot(data = base_completa,
       mapping = aes(x= GPS,color = IncomeGroup)) +
  geom_histogram(fill = "lightcyan3",bins = 100)+
  labs(title = "Histograma de Gasto Publico en Salud",y = "Frecuencia", caption = "DatosMacro")+
  theme_classic()
```

```
## Warning: Removed 2162 rows containing non-finite values (stat_bin).
```

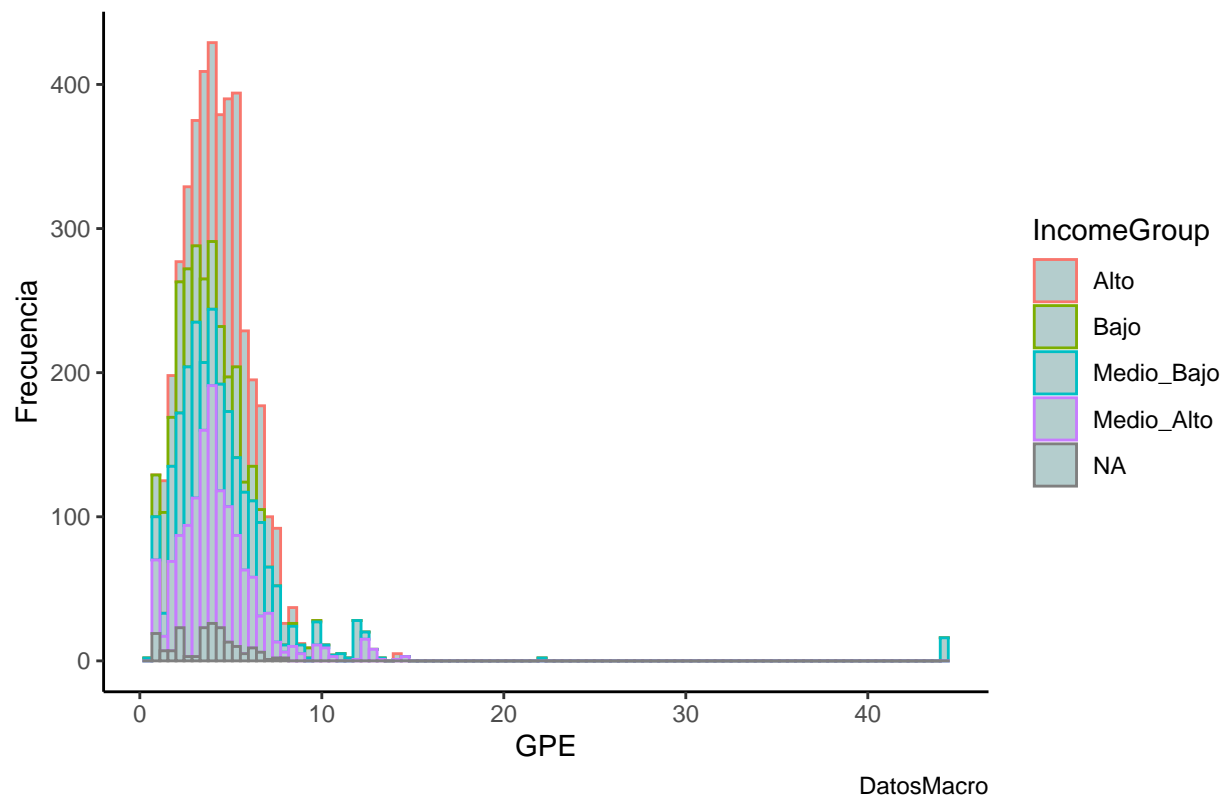
Histograma de Gasto Publico en Salud



```
## GPE
ggplot(data = base_completa,
       mapping = aes(x= GPE,color = IncomeGroup)) +
  geom_histogram( fill = "lightcyan3",bins = 100)+
  labs(title = "Histograma de Gasto Publico en Educacion",y = "Frecuencia", caption = "DatosMacro")+
  theme_classic()
```

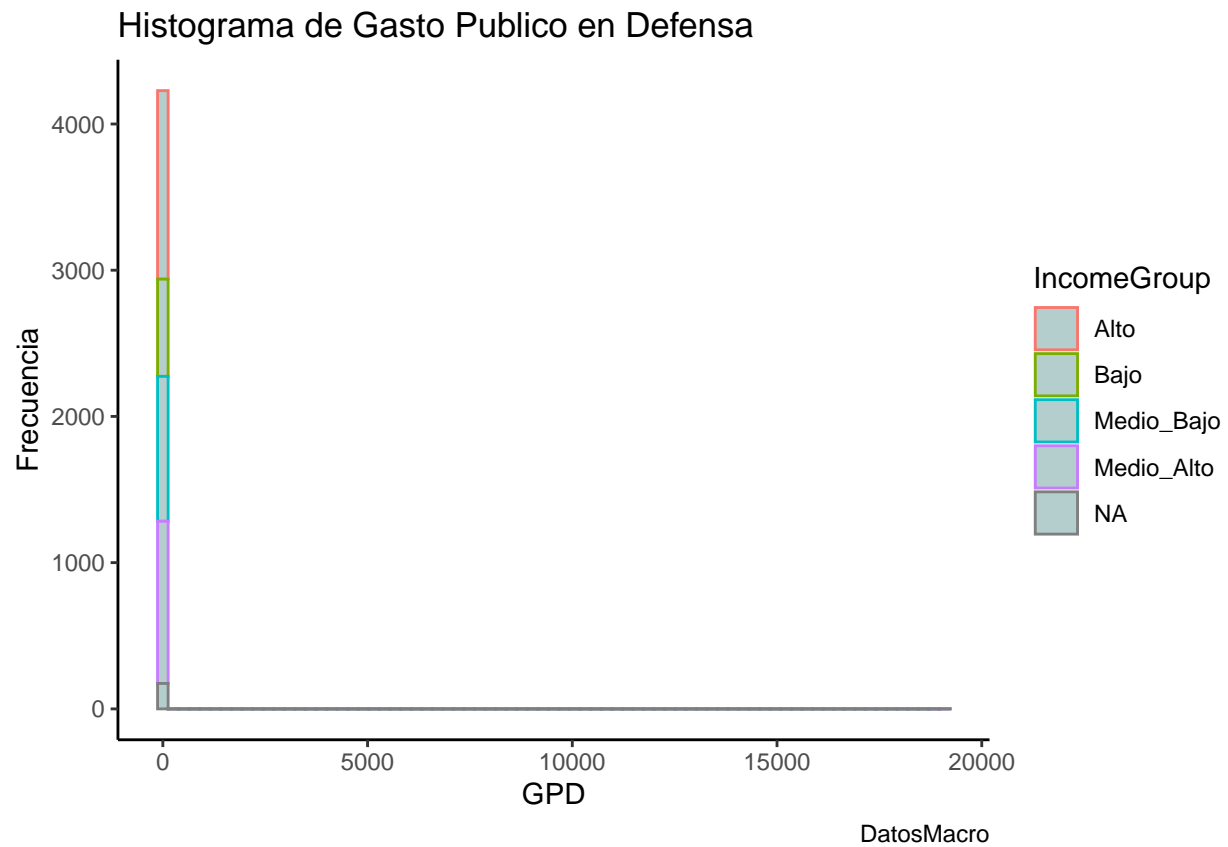
Warning: Removed 1965 rows containing non-finite values (stat_bin).

Histograma de Gasto Publico en Educacion

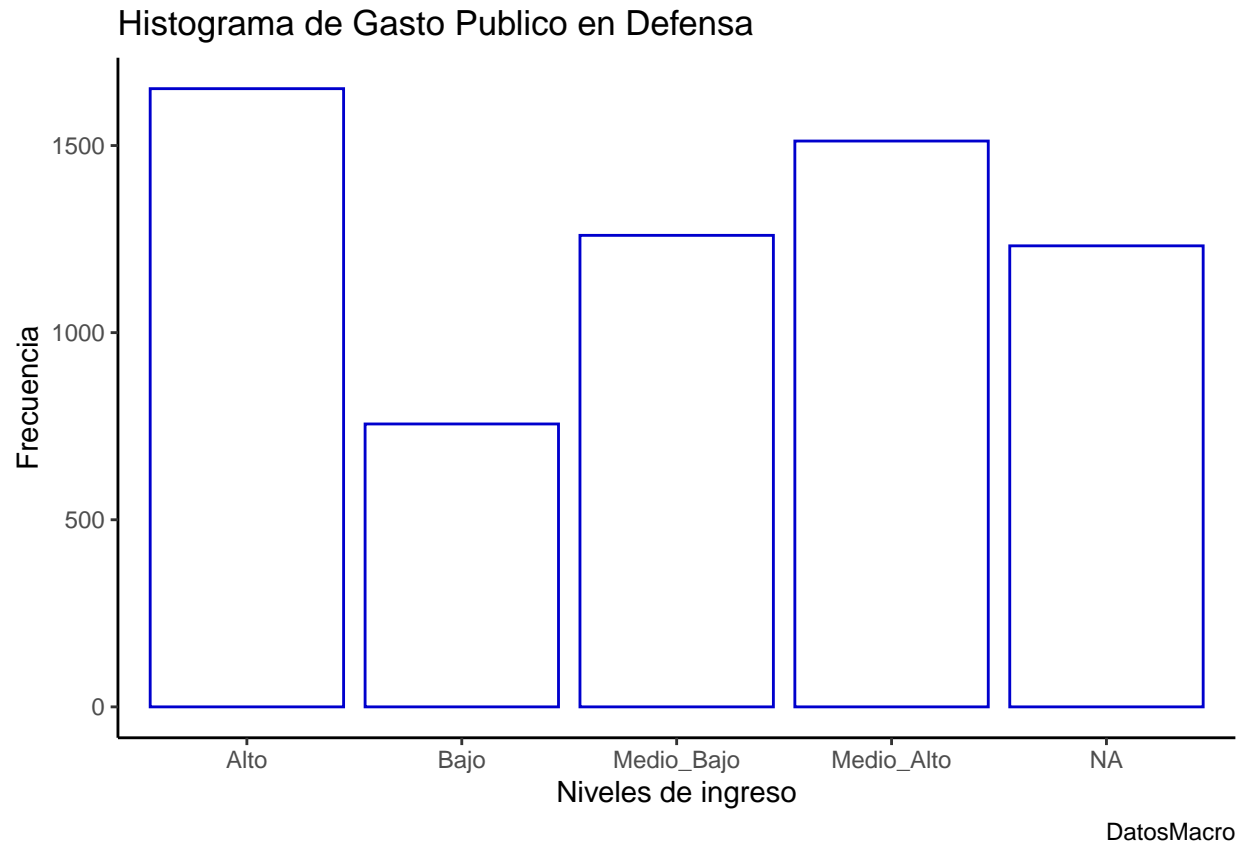


```
## GPD
ggplot(data = base_completa,
       mapping = aes(x= GPD,color = IncomeGroup)) +
  geom_histogram(fill = "lightcyan3",bins = 75)+
  labs(title = "Histograma de Gasto Publico en Defensa",y = "Frecuencia", caption = "DatosMacro")+
  theme_classic()
```

Warning: Removed 2183 rows containing non-finite values (stat_bin).



```
## IncomeGroup
ggplot(data = base_completa,
       mapping = aes(x = IncomeGroup)) +
  geom_bar(color = "mediumblue", fill = "white")+
  labs(title = "Histograma de Gasto Publico en Defensa", y = "Frecuencia", x = "Niveles de ingreso", caption = "DatosMacro")
  theme_classic()
```



Punto 7

```
base_completa<- tibble(base_completa)
```

```
class(base_completa)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
variables_cor <- base_completa %>% select(IHSM,GINId,IC,IDH,GPS,GPE,GPD)
```

```
cor(variables_cor, use = "complete.obs")
```

```
##           IHSM      GINId      IC      IDH      GPS      GPE
## IHSM    1.00000000 -0.2442584  0.73767861  0.73767861  0.4561231  0.38535472
## GINId -0.24425842  1.0000000 -0.42793150 -0.42793150 -0.1544074 -0.30498732
## IC     0.73767861 -0.4279315  1.00000000  1.00000000  0.4955931  0.46636303
## IDH    0.73767861 -0.4279315  1.00000000  1.00000000  0.4955931  0.46636303
## GPS    0.45612311 -0.1544074  0.49559314  0.49559314  1.0000000  0.27945615
## GPE    0.38535472 -0.3049873  0.46636303  0.46636303  0.2794561  1.00000000
## GPD    0.02206855  0.1029726 -0.06786109 -0.06786109 -0.1129043 -0.06978714
##           GPD
## IHSM    0.02206855
## GINId    0.10297255
```

```
## IC      -0.06786109
## IDH     -0.06786109
## GPS     -0.11290430
## GPE     -0.06978714
## GPD      1.00000000
```

Punto 8